

Improving Atmospheric Models by Accounting for Chaotic Physics

PRASHANT D. SARDESHMUKH,^{a,b} JIH-WANG AARON WANG,^{a,b} GILBERT P. COMPO,^{a,b} AND CÉCILE PENLAND^b

^a Cooperative Institute for Research in Environmental Sciences, University of Colorado Boulder, Boulder, Colorado

^b NOAA/ESRL/Physical Sciences Laboratory, Boulder, Colorado

(Manuscript received 23 November 2022, in final form 24 March 2023, accepted 1 May 2023)

ABSTRACT: It is well known that randomly perturbing an atmospheric model's diabatic tendencies can increase its probabilistic forecast skill, mainly by increasing the spread of ensemble forecasts and making it more consistent with the errors of ensemble-mean forecasts. Less obvious and less well established is that such perturbations can also reduce the errors of the ensemble-mean forecasts and improve the model's mean climate, variability, and sensitivity to forcing. A clear reduction in ensemble-mean forecast errors is demonstrated here in large ensembles of 15-day forecasts made with NOAA's Global Forecast System model. The nearly ubiquitous reduction around the globe, obtained throughout the forecast range, is interpreted as arising in effect from a modification of the model's deterministic evolution operator by a stochastic noise-induced drift. The effect is general in systems with state-dependent noise, and occurs even if the noise is not white. In the atmospheric context considered here, the effect is suggested to arise largely from noise-induced reductions of mechanical and thermal damping by chaotic boundary layer and cloud-radiative processes, which also tend to increase model sensitivity to forcing. The results presented here are consistent with many previous studies performed with models ranging from simple stochastically forced models to comprehensive global weather and climate models. They suggest that the diabatic interactions in most current global atmospheric models may not be sufficiently chaotic and this deficiency could be partly remedied by specifying additional stochastic terms. Using some empirical guidance in such specifications may be unavoidable, given the generally intractable complexity of the diabatic interactions.

KEYWORDS: Atmospheric circulation; Climate prediction; Numerical weather prediction/forecasting; Probability forecasts/models/distribution; Model errors; Model evaluation/performance


1. Introduction

Global atmospheric models have come a long way since their initial development in the 1960s, leading to much better weather forecasts and a better understanding of the general circulation of the atmosphere including its mean, variability, and sensitivity to forcing. Global atmospheric models are also an integral part of Earth system models used for understanding past climate variations and projecting future climate changes. Given their central role, it is essential to continue improving atmospheric models, whose errors continue to compromise weather and climate predictions. Our goal here is to advance a case that a substantially chaotic but not necessarily unresolved portion of atmospheric diabatic processes is likely underrepresented in most such models and could be accounted for approximately by specifying additional stochastic terms in their evolution equations. The arguments presented complement those in recent overview articles (Leutbecher et al. 2017; Berner et al. 2017; Strommen et al. 2019; Palmer 2019) and draw additionally upon extensive sets of experimental 15-day ensemble forecasts made by us with and without stochastic terms in the Global Forecast System (GFS) model of the National Oceanic and Atmospheric Administration (NOAA).

The robustness of our results, and especially of their interpretation based on basic features of stochastically forced systems, suggest their wide relevance in Earth system modeling.

Adding stochastic terms to a global atmospheric model's equations was originally motivated in weather prediction contexts, specifically by a desire to increase the spread (i.e., the standard deviation) of ensemble forecasts in order to make it more consistent with the errors of ensemble-mean forecasts, which were (and still are) too often larger than anticipated from the ensemble spread. Randomly perturbing a forecast model's diabatic tendencies was shown to be beneficial in this regard (Buizza et al. 1999; Palmer et al. 2009), a basic result confirmed in numerous subsequent studies (Leutbecher et al. 2017). Less obvious, and less well established, is that such random perturbations can also reduce the errors of the ensemble-mean forecasts. We show particularly clear reductions here in our 15-day ensemble forecast experiments. The reductions are nearly ubiquitous in several key atmospheric variables, and throughout the forecast range, including near the 15-day limit of daily weather predictability when initial conditions are nearly forgotten and forecast errors are nearly saturated. We interpret such improvements in ensemble-mean forecast skill as indicative of model improvement through a crude accounting of chaotic diabatic processes in the atmosphere by stochastic perturbations of the parameterized physical tendencies (SPPT) in the model.

In this context it is important to distinguish between model improvement and forecast improvement, although the two are obviously related. To some degree, probabilistic forecasts can be improved without changing a model by improving initial and/or boundary conditions, using larger forecast ensembles, and making

 Denotes content that is immediately available upon publication as open access.

Corresponding author: Prashant D. Sardeshmukh, prashant.d.sardeshmukh@noaa.gov

DOI: 10.1175/JCLI-D-22-0880.1

© 2023 American Meteorological Society. For information regarding reuse of this content and general copyright information, consult the [AMS Copyright Policy \(www.ametsoc.org/PUBSReuseLicenses\)](#).

a posteriori bias corrections based on past forecast error statistics. Less intuitively, model improvement is also distinct from improving a model's mean climate and variability through model tuning (including "flux-correction" and "nudging" approaches). Tuning can certainly improve some aspects of a model's climate, although one cannot be certain to what extent this is achieved by introducing additional compensating model errors instead of reducing the original model errors. By treating, in effect, the symptoms rather than the causes of the original errors, tuning can degrade other aspects of a model's climate. This is a concern especially in multi-component systems such as the atmosphere, in which a model error in one component can affect the behavior of another in just a few hours (Klinker and Sardeshmukh 1992; Rodwell and Palmer 2007).

To fix ideas and establish terminology, we define model improvement here as introducing tendency correction terms R in a model's evolution equations as

$$dx/dt = a(\mathbf{x}) + R, \quad (1)$$

where \mathbf{x} is the model state vector and $a(\mathbf{x})$ is its uncorrected tendency. One may think of specifying R as a sum of deterministic and random terms

$$R \approx R_d(\mathbf{x}) + \mathbf{B}(\mathbf{x})\boldsymbol{\eta}, \quad (2)$$

where $\mathbf{B}(\mathbf{x})$ is a matrix and $\boldsymbol{\eta}$ is a vector (not necessarily of the same dimension as \mathbf{x}) with independent random components η_m , each with mean 0, variance 1, and an autocorrelation time scale τ_m . The R_d term reflects an attempt to reduce the model tendency error at each model time step. By contrast, since $\boldsymbol{\eta}$ is random, the $\mathbf{B}(\mathbf{x})\boldsymbol{\eta}$ term can either decrease or increase the tendency error at individual time steps, which might make one question its corrective role. The answer is that it can improve the conditional (i.e., the forecast) as well as the unconditional (i.e., the "climate") probability distributions of \mathbf{x} including their mean, variance, skewness, and tails. As already noted, there is strong evidence that such random terms improve the forecast ensemble spread and some, albeit less clear-cut, evidence that they also improve the ensemble mean forecasts and the model's climate. The effect on the ensemble mean is evident even over short time intervals Δt , and can be made explicit, if $\eta_m \Delta t$ can be approximated as a random walk $\phi_m \sqrt{2\tau_m} \sqrt{\Delta t}$ associated with Gaussian white noise integrated over Δt , where ϕ_m is a standard Gaussian variable with mean 0 and variance 1 (appendix). Defining $\beta_{im} = B_{im} \sqrt{2\tau_m}$, the change Δx_i in each component x_i of \mathbf{x} over Δt may then be expressed (ignoring the R_d term for a moment) as

$$\Delta x_i = [a_i(\mathbf{x}) + D_i(\mathbf{x})]\Delta t + \sum_m \beta_{im}(\mathbf{x}) \phi_m \sqrt{\Delta t}, \quad (3)$$

where

$$D_i(\mathbf{x}) = \frac{1}{2} \sum_j \sum_m \frac{\partial \beta_{im}(\mathbf{x})}{\partial x_j} \beta_{jm}(\mathbf{x})$$

is the noise-induced drift (see appendix for a simple derivation), which directly affects the deterministic dynamics and

hence the ensemble mean, whereas the $\sqrt{\Delta t}$ term represents a Wiener process (i.e., a random walk) which directly affects the ensemble spread. This noise-induced drift D contributes to R_d in (1). It depends sensitively on $\partial \mathbf{B}/\partial \mathbf{x}$, whereas the effect of the noise on the spread depends more robustly on \mathbf{B} itself. Note however that even if \mathbf{B} is independent of \mathbf{x} (i.e., even if the noise is "additive" and not "multiplicative") it can still affect ensemble-mean forecasts at longer forecast ranges as well as the model's climate through formally second-order, but not necessarily small, nonlinear modifications of $a(\mathbf{x})$ associated with, say, altered resolved fluxes. We will find the Gaussian white noise approximation (3) to be useful for understanding the impacts of even the red noise perturbations specified in our experiments and also for distinguishing between multiplicative and additive noise effects.

Figure 1 gives a preview of the beneficial impacts of the random tendency perturbations specified in our GFS experiments on both the mean and spread of 15-day ensemble forecasts of several key atmospheric variables (upper-tropospheric vorticity, midtropospheric vertical velocity, near-surface air temperature, and column precipitable water). Results are shown for 80-member ensemble forecasts generated both with and without random perturbations for 80 distinct forecast cases in the January–March 2016 period. A more detailed presentation will be made in section 4. The random perturbations are introduced in the tendency equation for each model component x_i as

$$dx_i/dt = A_i + (1 + \mu_i r_i) P_i, \quad (4)$$

where $A_i(\mathbf{x})$ and $P_i(\mathbf{x})$ are the adiabatic and parameterized diabatic tendencies. Note that the vertical diffusion of momentum, temperature, and specific humidity is included in the diabatic tendency and is perturbed, whereas the horizontal diffusion of these variables is included in the adiabatic tendency and is not perturbed. The random numbers r_i are Gaussian deviates with zero mean and variance 0.8, and μ_i is a vertical weight function specified to be 1 everywhere except near highest and lowest model levels where it is tapered to 0. The r_i are constant in vertical columns and have specified horizontal and temporal correlation scales of $L_r = 500$ km and $\tau_r = 6$ h, respectively. Note that these scales are much larger than the spatial and temporal discretization of the model (grid size ~ 70 km; time step ~ 20 min for dynamics and ~ 10 min for physics), so the noise cannot be considered as accounting only for subgrid scale chaotic feedbacks on the resolved scales. Equation (4) is of the form of (1), as can be seen by writing $A_i(\mathbf{x}) + P_i(\mathbf{x}) = a_i(\mathbf{x})$ and

$$P_i(\mathbf{x}) \mu_i r_i = \sum_m P_i(\mathbf{x}) C_{im} \eta_m = [\mathbf{B}(\mathbf{x})\boldsymbol{\eta}]_i,$$

where \mathbf{C} is a constant matrix related to the spatial covariances of the random perturbations.

Consistent with numerous previous studies, Fig. 1 shows an increase in the forecast ensemble spread by introducing the random tendency perturbations. More unusually, it also shows a reduction in the ensemble-mean forecast error throughout the 15-day forecast range. Evidently, multiplying the parameterized diabatic tendency of each model variable at each grid

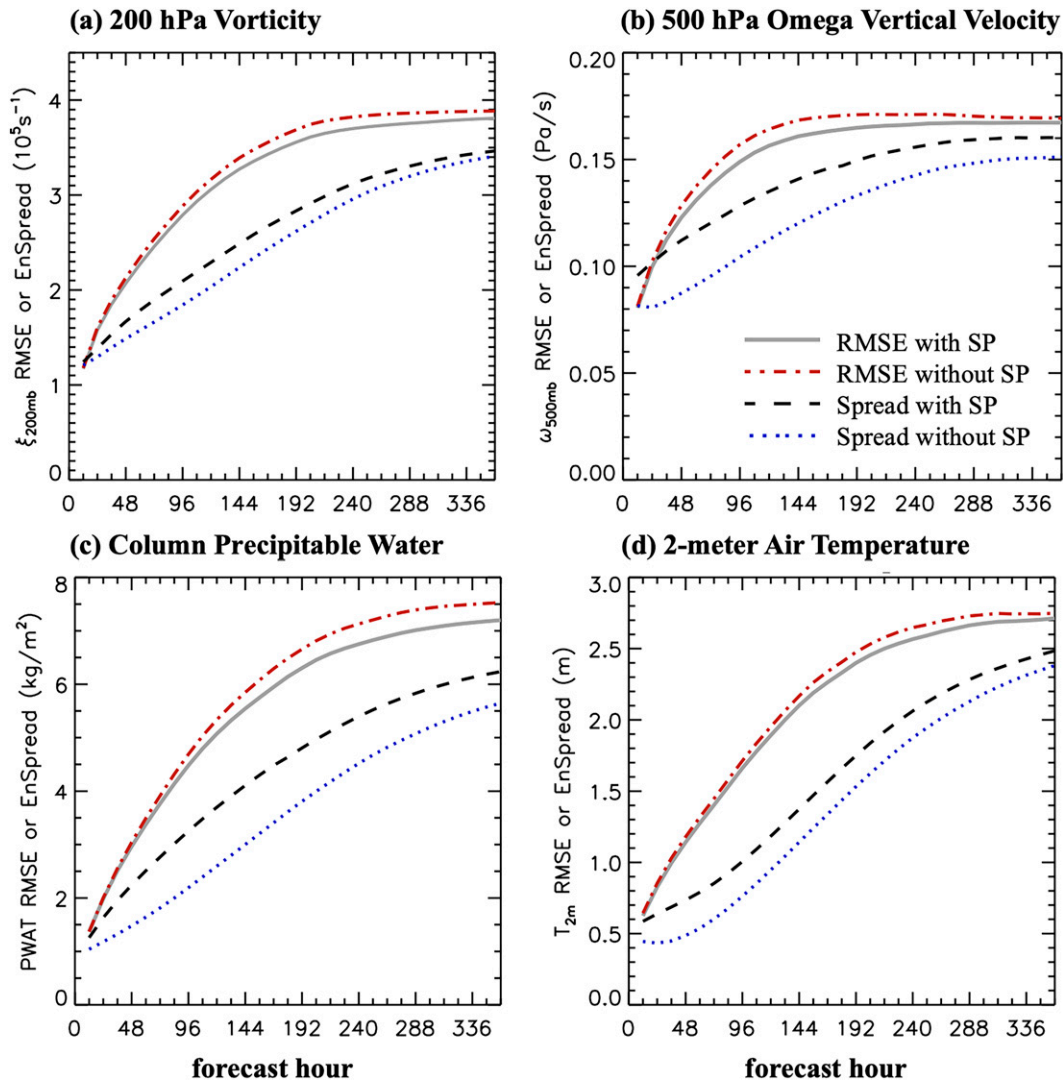


FIG. 1. Global root-mean-square error of ensemble-mean forecasts (RMSE) and forecast ensemble spread (spread) of 80-member ensemble forecasts from 80 different start times in the January–March 2016 period made using the GFS model. Results are shown for (a) 200-hPa vorticity, (b) 500-hPa vertical p -velocity omega, (c) column-integrated precipitable water, and (d) 2-m air temperature $T_{2\text{m}}$. In each of (a)–(d), the highest and lowest curves show, respectively, the RMSE and spread of the forecasts without stochastic parameterization in the model. When the stochastic parameterizations are included, the RMSE curves are lowered and the spread curves are raised, reducing the error/spread gap in each panel.

point at each time step by a random factor approximately between 0 and 2 not only does not degrade the ensemble-mean forecasts, but actually improves them. Understanding the basic reasons for this, and the implications for improving weather and climate models, are our chief concerns in this paper.

Section 2 discusses how the impacts of model changes on forecast error/spread relationships such as in Fig. 1 may be interpreted in the context of simple stochastically forced linear systems. Section 3 provides a perspective on how a change in stochastic forcing can affect the mean, variance, skewness, and tails of the stationary “climate” probability distribution of a system, again in the context of simple linear and nonlinear

systems for which analytical expressions can be derived. Section 4 provides more details of our GFS experiments, demonstrates the near-ubiquity of the beneficial impacts of the random perturbations, and assesses the relative importance of randomness in different aspects of the model physics. It also shows the impacts of the random perturbations on the non-Gaussian aspects of the forecast distributions such as their skewness and kurtosis, which are important for assessing extreme anomaly risks (Sardeshmukh et al. 2015). Section 5 attempts to gain further understanding of the GFS results by considering the impacts of stochastic mechanical and thermal damping, and a discussion and summary follow in section 6.

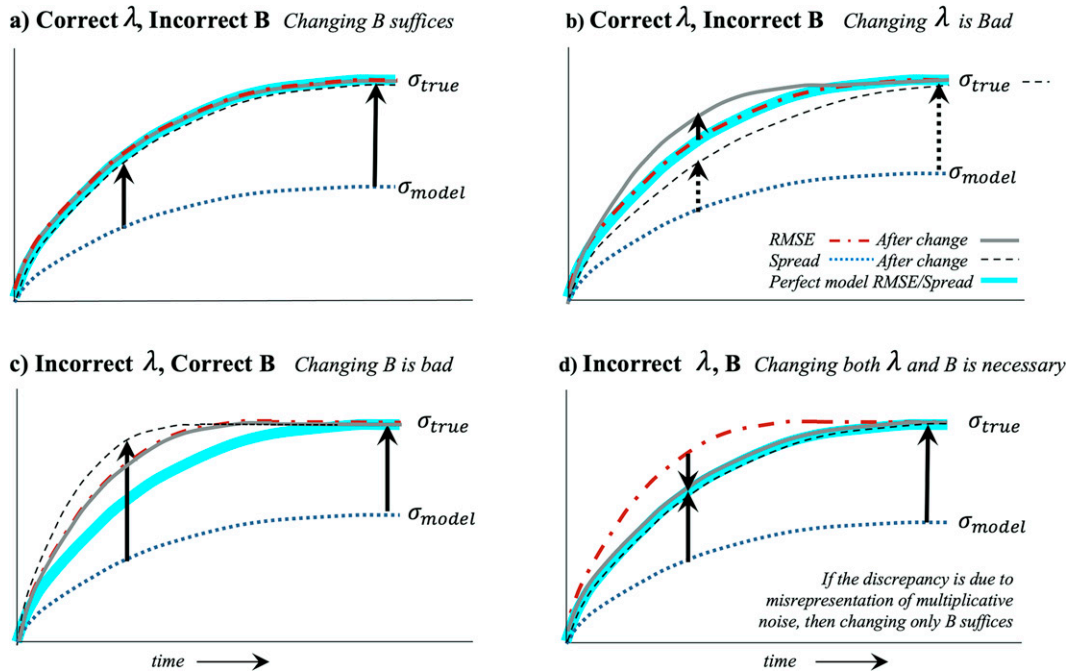


FIG. 2. (a)–(d) The problem with inferring model improvement from a reduction in the forecast error/spread gap. The figure illustrates the impact of changing (solid vertical segments) the damping rate λ and/or the stochastic forcing amplitude B of imperfect univariate linear Markov models on the RMSE, spread, and the error/spread gap in four different scenarios. For the perfect model, the RMSE and spread curves are identical, as indicated by the thick cyan curve in each panel. The figure shows that, even in this simplest model setting, a reduction in the error/spread gap from model changes is an ambiguous indicator of model improvement. The gap is reduced in all four scenarios but is associated with model improvement in only two of them [(a) and (d)], as discussed in the text.

2. Interpreting model improvement in a forecast error/spread framework

A gap between the ensemble-mean forecast error and ensemble spread growth curves compromises probabilistic predictions. Its reduction through model changes does not, however, always indicate model improvement. We argue below that to qualify unambiguously as model improvement, a smaller gap must also generally be associated with a reduction in the ensemble mean forecast error.

Consider the simplest univariate Markov system of the form (1),

$$dx/dt = -\lambda x + B\eta, \tag{5}$$

where x , λ , B , and η are scalars and η is Gaussian white noise. There are only two model parameters here, a linear damping rate λ and a stochastic forcing amplitude B , whose values for an imperfect model differ from the true values λ_{tr} and B_{tr} . The statistics of this system are Gaussian, with long-term means $\langle x \rangle = 0$ and long-term variances $\sigma^2 = B^2/(2\lambda)$ and $\sigma_{tr}^2 = B_{tr}^2/(2\lambda_{tr})$. An ensemble-mean forecast evolves from an initial condition x_{in} as

$$\bar{x} = \langle x(t)|x_{in} \rangle = e^{-\lambda t} x_{in}. \tag{6}$$

It can then be shown that the squared ensemble-mean forecast error $\langle \varepsilon^2 \rangle = \langle (\bar{x} - x_{obs})^2 \rangle$ and ensemble spread $\langle s^2 \rangle = \langle (x - \bar{x})^2 \rangle$ averaged over all initial conditions evolve as

$$\langle \varepsilon^2 \rangle(t) = \sigma_{tr}^2 [1 + e^{-2\lambda t} - 2e^{-(\lambda_{tr} + \lambda)t}] \quad \text{and} \tag{7}$$

$$\langle s^2 \rangle(t) = \sigma^2 (1 - e^{-2\lambda t}). \tag{8}$$

These curves are identical for a perfect model with $\lambda = \lambda_{tr}$, $B = B_{tr}$, and $\sigma^2 = \sigma_{tr}^2$. One does not know them if one does not know λ_{tr} and B_{tr} , but one does know their asymptotic value $\sigma_{tr}^2 = \sigma_{obs}^2$ from the system observations x_{obs} . One can also use those observations to determine the forecast error curve directly for an imperfect model, without knowing the right-hand side of (7). Note that this error curve always lies above the perfect model error curve if $\lambda \neq \lambda_{tr}$. We consider the question of what can be inferred about model improvement from model changes given only the true σ_{tr}^2 and the original and modified forecast error and spread curves.

Figure 2 illustrates scenarios for imperfect models in which either λ or B or both differ from their true values, so the gap between the error and spread curves is not zero. For brevity we focus on models for which σ^2 is smaller than σ_{tr}^2 . To close the error/spread gap, the two must be made equal. One may accomplish this by increasing B , decreasing λ , or doing both. But would all of these actions result in model improvement, that is, bring the model's λ and B closer to λ_{tr} and B_{tr} ? Figure 2 considers four possibilities. In Fig. 2a, $\lambda = \lambda_{tr}$ and $B \neq B_{tr}$. In this case the error curve is already identical to the perfect model

error curve (although one does not know this), and increasing B to make $\sigma^2 = \sigma_{tr}^2$ would make the spread curve identical to the error curve. One would then know that one has arrived at a perfect model. If, however, one were to change λ but not B to make $\sigma^2 = \sigma_{tr}^2$, as in Fig. 2b, one would still reduce the error/spread gap but not be able to close it because λ would now be incorrect. In Fig. 2c, $\lambda \neq \lambda_{tr}$ and $B = B_{tr}$. In this case changing B to make $\sigma^2 = \sigma_{tr}^2$ would also reduce the error/spread gap, but one would not be able to close it because both λ and B would now be incorrect. Figure 2d considers the most general possibility, that both $\lambda \neq \lambda_{tr}$ and $B \neq B_{tr}$. If one suspects this to be the case, one would consider changing both to make $\sigma^2 = \sigma_{tr}^2$, but this could clearly be done in an infinite number of ways. Only some of them would reduce the error (by bringing λ closer to λ_{tr}), and only one would result in closing the gap at all forecast ranges.

Even this simple thought experiment highlights the fact that a reduction in the forecast error/spread gap is by itself an ambiguous indicator of model improvement. The gap is reduced in all four cases considered in Fig. 2, but is associated with model improvement in only two of them (Figs. 2a,d). In particular, a reduction in the gap obtained by only increasing the spread does not unambiguously indicate model improvement: it is associated with an improved model in one case (Fig. 2a) but a degraded one in another (Fig. 2c). One might regard the scenarios depicted in Figs. 2a–c as unlikely, since they are associated with either λ or B already being perfect. In the more likely scenario of Fig. 2d in which both λ and B are imperfect, reductions in both the gap and ensemble-mean forecast error would be necessary to establish model improvement. Introducing stochastic tendency perturbations in the GFS model evidently satisfies these two conditions in Fig. 1. In the context of Fig. 2d, the effect of the stochastic tendency perturbations in our GFS experiments is not only to increase the spread by effectively increasing B , but also to reduce the error by effectively correcting λ through the noise-induced drift in (3) [or more transparently, through (18) in section 5].

Note that the long-term mean “climate” response of the model (5) to a steady forcing F is $\langle x \rangle = F/\lambda$, which is incorrect if $\lambda \neq \lambda_{tr}$. A reduction in the ensemble-mean forecast error, which can only be achieved by bringing an imperfect model’s λ closer to λ_{tr} , would thus also imply an improvement in the model’s mean climate and sensitivity to external forcing.

3. The basic effects of additional stochastic forcing

It is notable in Fig. 1 that the reductions in both the ensemble-mean forecast error and the error/spread gap occur throughout the 15-day forecast range, including beyond day 10 when the error and spread have begun to saturate. This suggests a beneficial impact of the stochastic perturbations on even the long-term climate statistics of the model. As in the previous section, one can gain some insight into this asymptotic behavior by considering the impact on the stationary probability distributions of simple dissipative linear and nonlinear 1D systems of the form (1),

$$dx/dt = a(x) + B(x)\eta, \tag{9}$$

where x , a , B , and η are scalars and η is Gaussian white noise. The Fokker–Planck equation governing the evolution of the probability density function $p(x|x_{in})$ from an initial condition x_{in} is

$$\frac{\partial p}{\partial t} = -\frac{\partial}{\partial x} \left[\left(a + \frac{1}{4} \frac{\partial B^2}{\partial x} \right) p \right] + \frac{1}{2} \frac{\partial^2}{\partial x^2} (B^2 p) \tag{10}$$

in which the noise-induced drift $(1/4)\partial B^2/\partial x$ has been explicitly included. The stationary probability density $p(x)$ is obtained by setting the time derivative to zero and rearranging to give

$$\frac{\partial}{\partial x} (B^2 p^2) = 4p^2 a \tag{11}$$

whose solution is

$$p(x) = p(x_0) \left[\frac{B^2(x_0)}{B^2(x)} \right]^{1/2} \exp \int_{x_0}^x \frac{2a(x')}{B^2(x')} dx' \tag{12}$$

in which the arbitrary x_0 may be chosen to be the location of the global maximum of $p(x)$, where $a = (1/4)\partial B^2/\partial x$.

It is important to recognize that although $a(x)$ and $B^2(x)$ uniquely determine $p(x)$, the converse is not true. Indeed, an infinite number of physically meaningful $[a(x), B^2(x)]$ function pairs can be consistent with the same $p(x)$. As is clear from (11), for any density function $p(x)$ one can solve for $B^2(x)$ given $a(x)$, and vice versa. This means, for example, that a nonlinear multiplicative noise-driven system $[a(x), B^2(x)]$ has the same $p(x)$ as a nonlinear additive noise-driven system

$$[\tilde{a}(x), b^2] = \left\{ \left[a(x) - \frac{1}{4} \frac{\partial B^2(x)}{\partial x} \right] \frac{b^2}{B^2(x)}, b^2 \right\}$$

with a constant b^2 . Resolving such ambiguities is important for model improvement, that is, for bringing a model’s erroneous $[a(x), B^2(x)]$ pairing closer to the “true” pairing.

Increasing the stochastic forcing by a factor $\gamma \geq 1$ to $\tilde{B}^2 = \gamma B^2$ changes the stationary $p(x)$ in (12) to

$$\tilde{p}(x) = \tilde{p}(x_0) \left[\frac{p(x)}{p(x_0)} \right]^{1/\gamma} \left[\frac{B^2(x_0)}{B^2(x)} \right]^{(\gamma-1)/(2\gamma)}. \tag{13}$$

If the exponential factor on the right-hand side on (12) is less than 1, as it is in dissipative systems in which $a(x_0) = 0$, then it can be shown that the pdfs in (13) satisfy the relationships

$$0 \leq \frac{p(x)}{p(x_0)} \leq \frac{\tilde{p}(x)}{\tilde{p}(x_0)} \leq 1 \quad \text{and} \tag{14a}$$

$$\frac{\tilde{p}(x_0)}{p(x_0)} \leq \frac{\tilde{p}(x)}{p(x)}. \tag{14b}$$

Relationship (14a), together with the fact that the integrals of both $p(x)$ and $\tilde{p}(x)$ over all x are 1, implies that $\tilde{p}(x)$ is

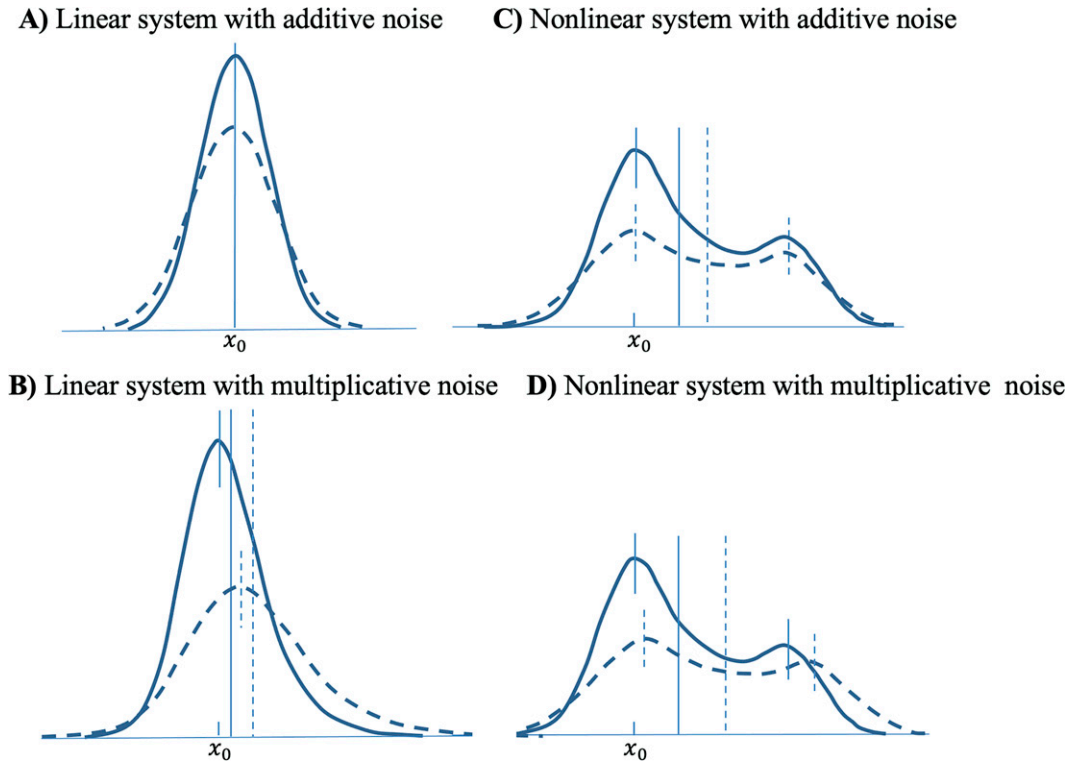


FIG. 3. (a)–(d) The impact of increasing the magnitude of the stochastic forcing in four different types of univariate systems on their stationary probability density function (pdf). In each of (a)–(d), the solid and dashed curves show the pdf before and after the increase in the forcing. The long solid and dashed vertical lines show the pdf mean, and the short solid and dashed vertical segments show the pdf maxima, before and after the forcing increase. Note that the original pdfs are identical in systems C and D as shown in (c) and (d), respectively. The figure shows that the variance increases in all four systems, but the impacts in the mean, skewness, and heaviness of the tails depend on whether the noise forcing is additive or multiplicative, as discussed in the text.

smoother and broader than $p(x)$, and relationship (14b) implies that this smoothing reduces the peak density at x_0 by a larger factor than at any other x .

Figure 3 schematically illustrates the impact of $\gamma \geq 1$ on the $p(x)$ of systems with four different types of $[a(x), B^2(x)]$ pairing, namely,

A—a linear system with additive noise:

$$[-a_0x, b^2], \quad (15a)$$

B—a linear system with multiplicative noise:

$$[-a_0x, (Ex + G)^2 + b^2], \quad (15b)$$

C—a nonlinear system with additive noise:

$$[\hat{a}(x), b^2], \quad \text{and} \quad (15c)$$

D—a nonlinear system with multiplicative noise:

$$[a(x), B^2(x)], \quad (15d)$$

where a_0, E, G , and b^2 are positive constants. System A, which is the same as in the previous section, is linearly damped and forced with additive noise. System B is also linearly damped,

but is forced with multiplicative noise. The specific form of this multiplicative noise is consistent with a combination of correlated additive and multiplicative (CAM) noises, whose relevance in understanding observed skewed and heavy-tailed atmospheric probability distributions was highlighted by Sardeshmukh and Sura (2009), Sardeshmukh and Penland (2015), and Sardeshmukh et al. (2015). Systems C and D are dissipative nonlinear systems forced by additive and multiplicative noises, respectively, and have the same stationary $p(x)$.

Increasing the stochastic forcing increases the variance in all four systems. In the simplest system A, however, this is all that happens: there is no change in the mean or shape (skewness, kurtosis, etc.) of the distribution as in the other three systems. The mean change in the other three systems is in the direction of the original skew, that is, positive in the cases illustrated. The skewness increases in systems B and D with multiplicative noise, but *decreases* in system C with additive noise. This contrast is interesting, and important for distinguishing between system C and system D, considering that both have the same original $p(x)$. Consistent with (13) and (14), the altered $\tilde{p}(x)$ in system C is more symmetric about the new mean than $p(x)$ is about the old mean, and is therefore less skewed, whereas in system D the altered $\tilde{p}(x)$ is more

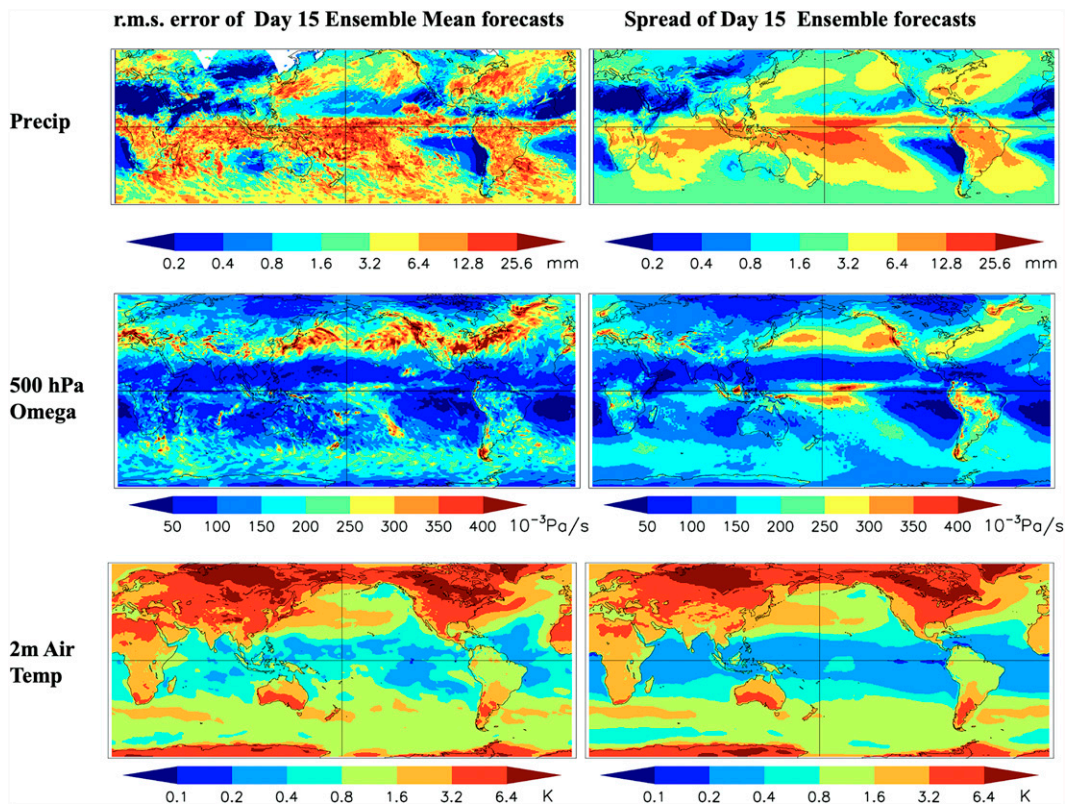


FIG. 4. (left) RMSE and (right) spread of the day-15 ensemble forecasts of (top) precipitation (12-h accumulations), (middle) 500-hPa vertical p -velocity omega, and (bottom) 2-m air temperature in the Control GFS forecast ensembles, which include the stochastic parameterizations in the model.

skewed. The basic reason for this increased skew (also in the CAM-noise-driven system B) is that the relatively larger $B^2(x)$ at large values of x makes the exponential factor in (12) larger (i.e., closer to 1). Because of this, the mean shift in system D is also larger than in system C.

The main message of Fig. 3 is that although increasing the stochastic forcing of a system “obviously” increases the variance of its stationary probability distribution, the impacts on the mean, shape, and tails of the distribution are more subtle and depend on whether the stochastic forcing is additive or multiplicative. The $P(x)r$ tendency perturbations specified in (4) in our GFS experiments are clearly multiplicative noises. They affect the ensemble-mean forecasts throughout the forecast range as well as the long-term mean due to noise-induced drifts. They also increase the magnitude of the skewness and kurtosis of the probability distributions in the GFS experiments, as we show in the next section.

4. Impact of stochastic forcing in the GFS model

The details of our experimental setup and GFS model used are given in two recent papers (Wang et al. 2019; Wang and Sardeshmukh 2021) and will not be repeated here. Briefly, we used a model version with horizontal and vertical resolutions of T254 (~ 70 km) and 64 levels, respectively, and performed

15-day 80-member ensemble forecasts from 80 different start times in the January to March 2016 period. Such ensemble forecasts were generated both with and without stochastic perturbations. As described in Wang et al. (2019), the Control forecasts included, in addition to the stochastic tendency perturbations (SPPT), stochastic perturbations of the boundary layer specific humidity itself (SHUM) at each model time step that may be regarded as perturbations to the average humidity tendency over the time step. Wang et al. (2019) ran the forecasts out to only 7 days and did not consider the ensemble spread. Wang and Sardeshmukh (2021) extended the forecasts to 10 days, but they restricted their focus to the 200-hPa kinetic energy spectrum.

Figure 4 shows maps of the day-15 ensemble-mean forecast error and ensemble spread for precipitation rate (Precip), midtropospheric vertical pressure velocity omega (ω_{500}), and near-surface air temperature (T_{2m}) in the Control forecasts. The ω_{500} and T_{2m} errors were defined with respect to reanalyses performed using the same GFS model in the data assimilation algorithm (Wang et al. 2019), and the Precip errors with respect to the NASA/GPM observations (Huffman et al. 2014). The reanalyses and observations have their own errors, of course, and the ensemble-mean forecast error maps are also derived from a smaller sample of 80 day-15 ensemble-mean forecasts for the 80 forecast cases than the large sample

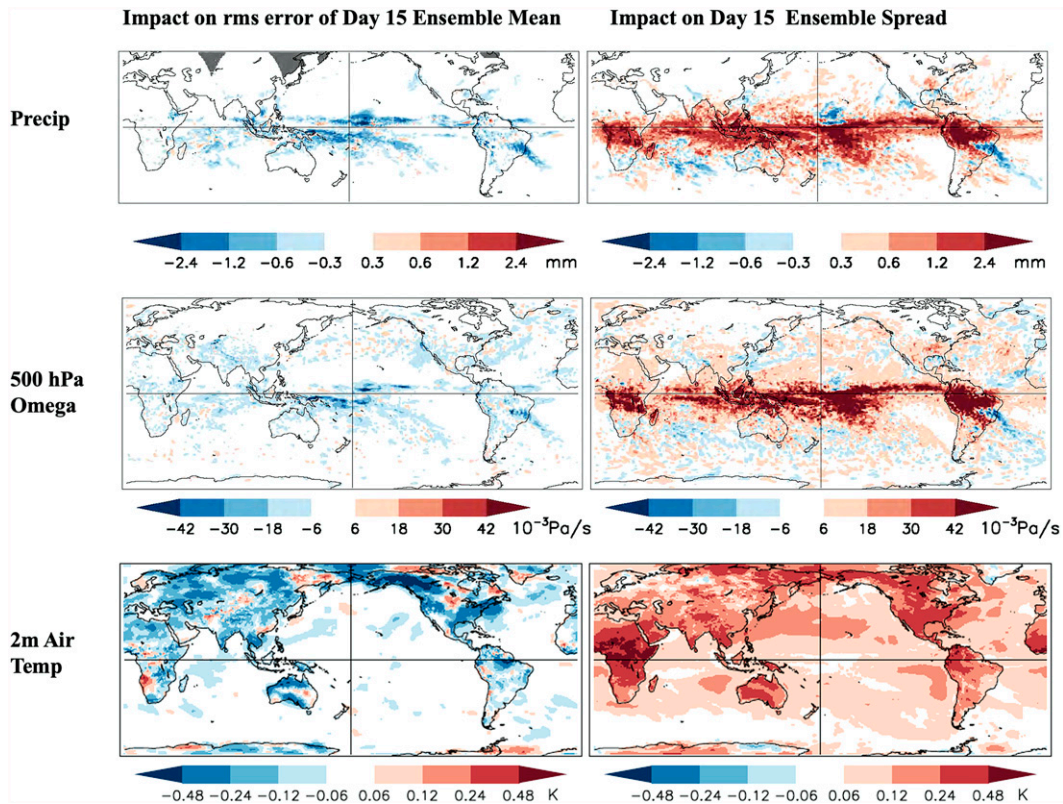


FIG. 5. As in Fig. 4, but showing the impact of the stochastic parameterizations on the day-15 RMSE and spread.

of 6400 ($=80 \times 80$) individual ensemble member forecasts used to construct the ensemble spread maps. Despite this, the error and spreads maps are reassuringly similar, although a general tendency for the error to be larger than the spread even after including the stochastic perturbations is evident, as was also seen in Fig. 1.

Figure 5 shows the impacts of the stochastic perturbations on the day-15 error and spread, determined by subtracting from Fig. 4 similar maps constructed from the parallel set of ensemble forecasts made without the stochastic perturbations. It shows relatively large increases in the spread of Precip and ω_{500} in the tropics and of T_{2m} over the continents. These are almost ubiquitously also associated with decreases of the ensemble-mean forecast error.

The global reductions in the error/spread gap and ensemble-mean errors shown in Fig. 1 are thus also associated with local reductions. Interestingly, the error is reduced in even some areas of *decreased* spread of Precip and ω_{500} , such as at 10°N just east of the date line, in the region extending south-eastward from central Brazil, and in the Gulf of Mexico.

These stochastic impacts are already evident early in the forecasts, as illustrated in Fig. 6 for 200-hPa vorticity at day 1 and day 5. This circulation variable is of particular interest, since its direct stochastic forcing, associated with the randomness only in the weak vertical diffusion in the free atmosphere above the boundary layer, is very small in the control forecasts. The day-1 maps for vorticity closely resemble those for

ω_{500} and Precip (not shown) and show that the impacts on the spread and error occur earlier in the tropics than in middle and high latitudes. While this does not prove that the extratropical impacts are “induced” by the tropical impacts, there is a hint that the impacts on the midlatitude spread are relatively strong near the Pacific and Atlantic jet entrances, consistent with noisy upper-tropospheric Rossby wave sources associated with perturbations to the tropical deep convection farther south (Sardeshmukh and Hoskins 1988). The stochastic tropical influence on extratropical spread was more cleanly demonstrated in Sardeshmukh (2005) in a numerical experiment in which a GCM was stochastically perturbed only in the tropics.

It is remarkable how the beneficial stochastic impacts on the forecast error and spread are evident locally, without any space-time filtering, even for variables with relatively small horizontal correlation scales such as precipitation, vertical velocity, and vorticity that are harder to predict than larger-scale variables such as 850-hPa temperature (T850), zonal wind (U850), or 500-hPa geopotential height (Z500). The beneficial impacts of stochastic parameterizations on T850, U850, and Z500 have been reported previously (e.g., Palmer et al. 2009; Zhu et al. 2018). We obtain them in our GFS experiments as well, as illustrated for T850 and Z500 at day 9 in Fig. 7. It is interesting to see the impacts on Z500 even in the tropics, where the anomalies of Z500 are very small and spatially almost uniform.

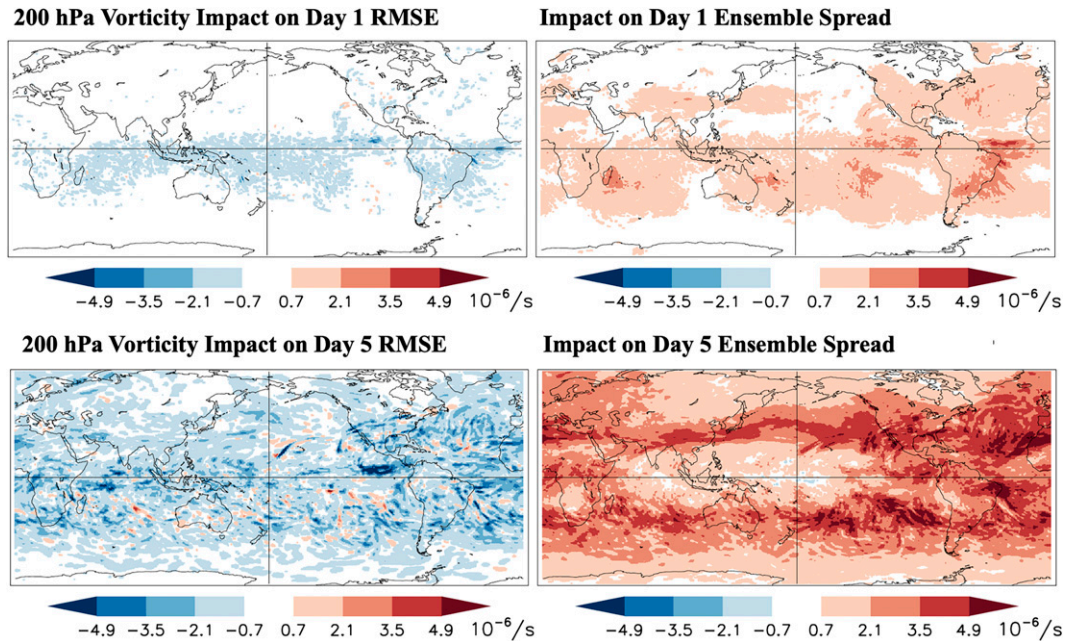


FIG. 6. As in Fig. 5, but showing the impacts of the stochastic parameterizations on the RMSE and spread of the 200-hPa vorticity at (top) day 1 and (bottom) day 5.

The stochastic perturbations also affect the skewness $S = \langle (x - \bar{x})^3 \rangle / \sigma^3$ and excess kurtosis $K = \langle (x - \bar{x})^4 \rangle / \sigma^4 - 3$ of the forecast ensembles, where \bar{x} and σ are the forecast ensemble mean and spread. Figure 8 shows maps of S and K of Precip in the day-15 control forecasts and how they are impacted by

the stochastic perturbations. Note that S and K are positive everywhere, attesting to the skewed and heavy-tailed character of precipitation distributions, and closely satisfy the $K = 1.5S^2$ property of gamma (Pearson type III) distributions that are often a good fit to observational precipitation data. The

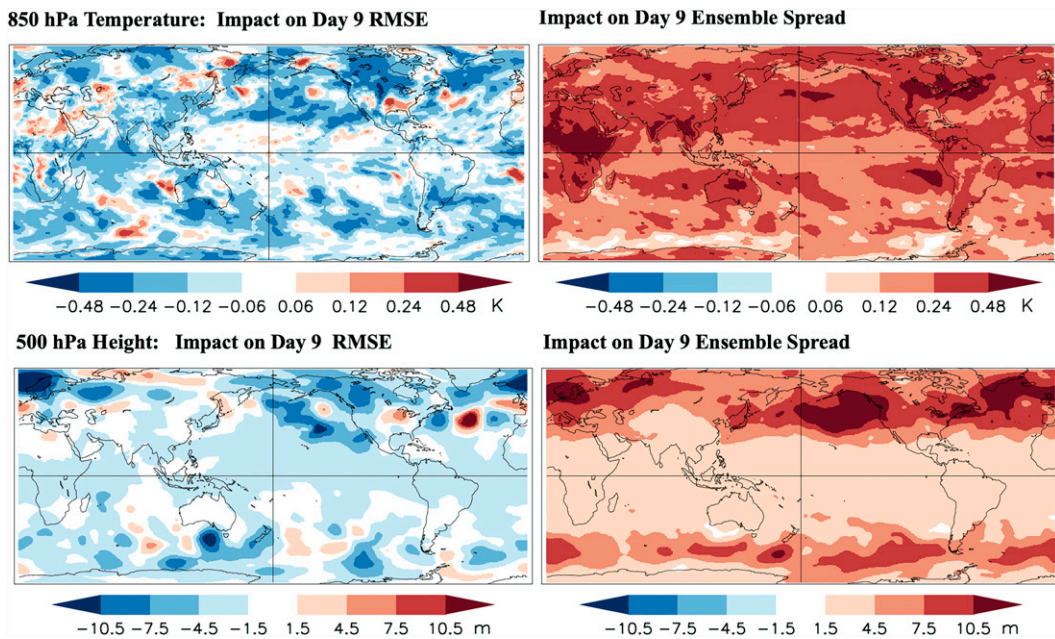


FIG. 7. As in Fig. 5, but for day-9 RMSE and spread of (top) 850-hPa temperature and (bottom) 500-hPa geopotential height.

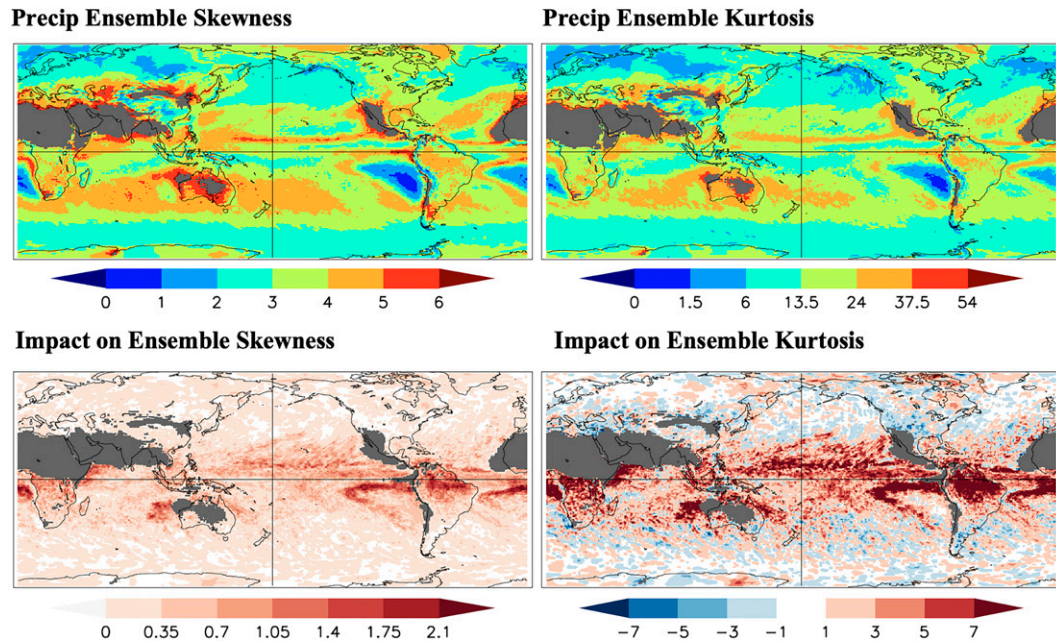


FIG. 8. (top) Forecast ensemble skewness S and excess kurtosis K of precipitation at day 15 in the GFS Control ensembles. Results are not shown in the driest desert regions of high sampling uncertainty. Note that S and K for gamma pdfs satisfy the relationship $K = 1.5S^2$, and the shading intervals are chosen to correspond to that relationship. The similarity of the left and right panels thus shows that precipitation pdfs are approximately gamma distributions nearly everywhere on the globe. (bottom) The impact of the stochastic parameterizations on S and K , which is to increase the skew everywhere and to increase the heaviness of the pdf tails in the tropics and subtropics associated with the increases of K in those regions.

substantial increases in the magnitudes of S and K in regions of large S and K are consistent with this property, and help satisfy it better in the control forecasts.

Figures 9 and 10 show similar results for ω_{500} and for T_{2m} , respectively, in a similar format to Fig. 8. Note that $\omega_{500} \approx -\rho g W_{500}$ is negative for upward motion, so the generally negative skewness of ω_{500} implies positive skewness of W_{500} . The stochastic perturbations increase the magnitudes of S and K of ω_{500} almost everywhere in regions of large S and K . The skewness of T_{2m} shows a more complicated spatial structure in Fig. 10, but the magnitudes of S and K are also generally increased in regions of large S and K , although not as clearly as in Fig. 9. Notably, the magnitudes of S and K are reduced in some regions such as the northwest Pacific, Hudson Bay, parts of the South Pacific convergence zone (SPCZ), and the northern Andes.

Just as for the ensemble-mean forecast error and ensemble spread, the stochastic impacts on S and K are also clearly evident early in the forecasts, as shown at day 5 for ω_{500} in Fig. 11. They are very similar to those at day 15 shown in Figs. 5 and 9, and are indeed somewhat larger, consistent with the argument made by Penland and Sardeshmukh (2022) that such impacts on forecast probability distributions are an inevitable feature of systems perturbed by multiplicative noise.

We also attempted to assess the relative importance of stochasticity in different aspects of the model physics by successively switching off the perturbations to the parameterized

vertical diffusion, convection, and radiation tendencies and the humidity in the Control forecasts.

Specifically, we generated three additional sets of 80-member forecast ensembles by switching off 1) the stochastic vertical diffusion tendencies (NO-VDSP), 2) the stochastic vertical diffusion + convection + radiation tendencies (NO-SPPT), and 3) the stochastic SPPT + specific humidity (NO-SP). Figure 12 summarizes the results from these denial experiments in a similar format to Fig. 1. It shows that the stochastic perturbations in all these processes make comparable contributions to increasing the spread, decreasing the error, and reducing the spread/error gap. The increased spreads of the thermodynamic variables T_{2m} and column precipitable water resulting from perturbing the (mostly boundary layer) vertical diffusion and specific humidity, respectively, are perhaps not surprising. Less obviously, these perturbations also increase the spreads of the circulation variables ω_{500} and 200-hPa vorticity and reduce the ensemble-mean forecast error of all four variables in Fig. 12. It is also evident that a large part of the total effect of SPPT on both the error and spread is due to stochastic perturbations of the vertical diffusion.

The additional dashed curve in all four panels in Fig. 12 shows that an almost identical reduction of RMSE by the stochastic perturbations is obtained in the control forecasts if only 40 of 80 members are used to calculate the ensemble mean. In other words, the noise-induced drift is large enough even at day 15 (when the ensemble mean anomaly is small as

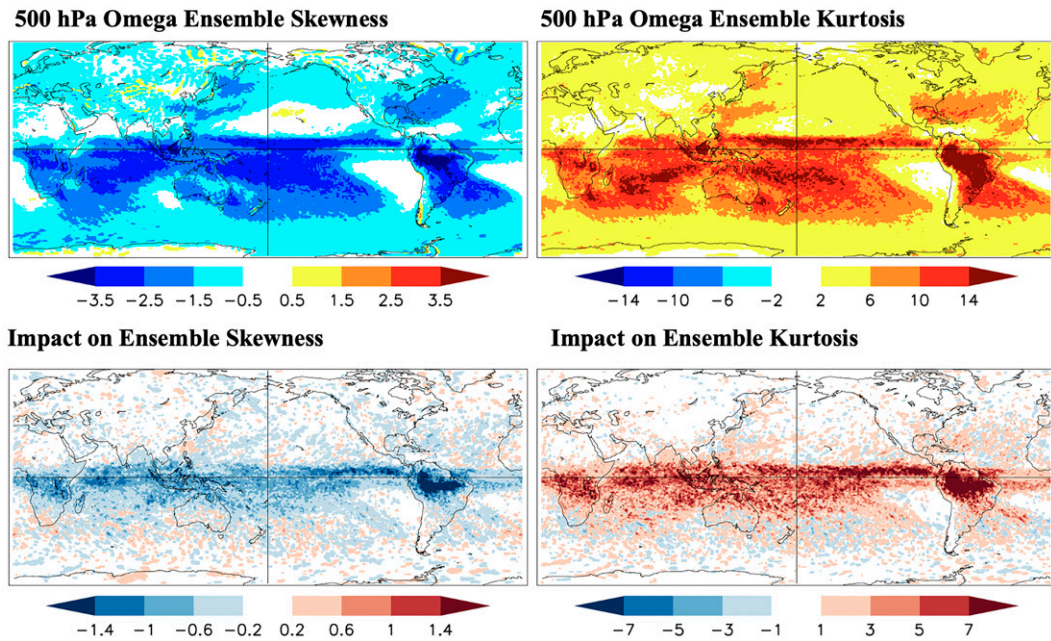


FIG. 9. As in Fig. 8, but for the day-15 ensemble skewness S and excess kurtosis K of the 500-hPa vertical p -velocity ω_{500} .

the ensemble mean forecasts converge to climatology) that its impact can be reasonably well estimated using only 40 members. This evidence that 40-member ensemble forecasts with stochastic perturbations can be superior to 80-member ensemble forecasts without stochastic perturbations has obvious practical implications for reducing the computational costs of ensemble forecasting.

We end this section noting that the impacts of stochastically perturbing the diabatic physics in GCMs are well known to be sensitive to the specified correlation length and time scales L , and τ , of the perturbations and are generally found to be smaller if these scales are reduced. Figure 13 illustrates this by contrasting the impacts on the day-15 error and spread of 200-hPa vorticity obtained using the default value

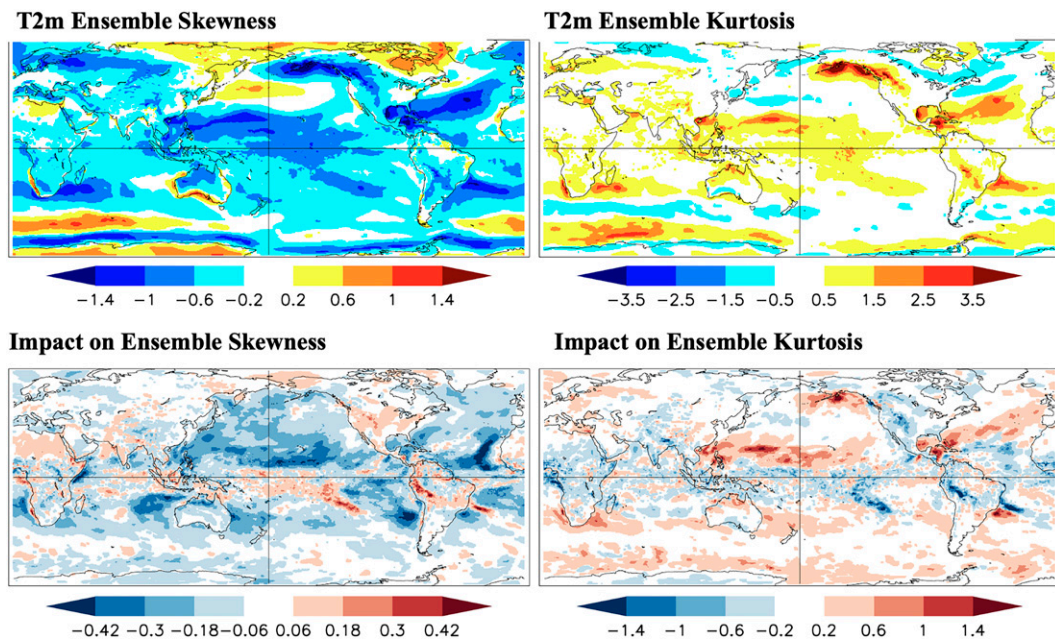


FIG. 10. As in Fig. 9, but for 2-m air temperature T_{2m} .

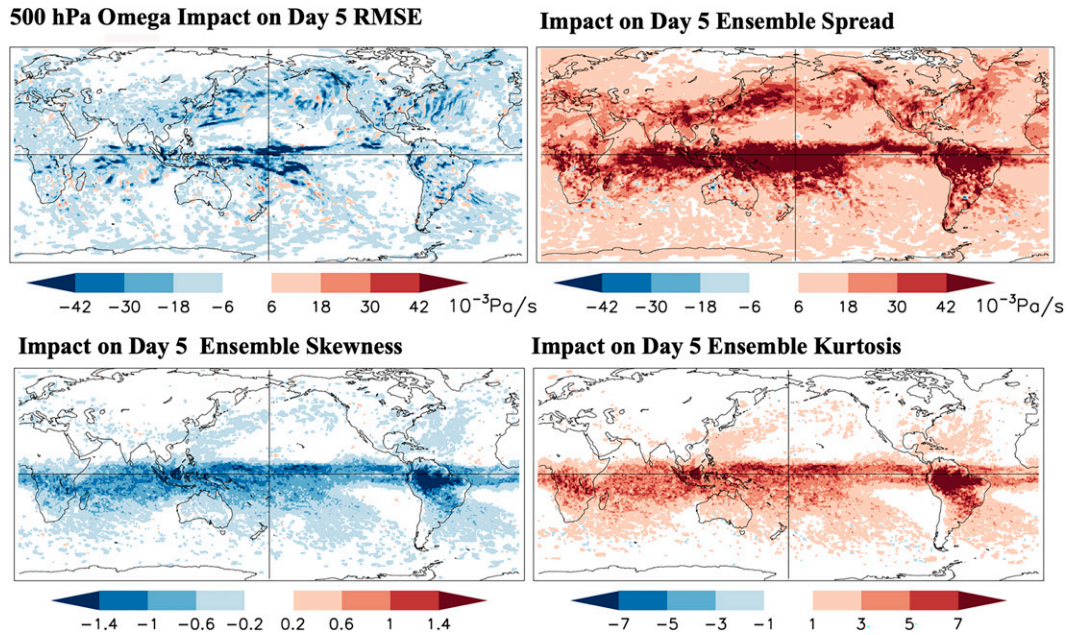


FIG. 11. (top) As in the middle row of Fig. 5, but showing the impact of the stochastic perturbations on the RMSE and spread of ω_{500} at day 5. (bottom) As in the bottom row of Fig. 9, but showing the impact of the stochastic perturbations on the skewness S and excess kurtosis K of ω_{500} at day 5.

of $L_r = 500$ km in our Control forecasts with those obtained using $L_r = 100$ km in another ensemble of otherwise identical forecasts. Similar weaker impacts using the smaller L_r were also obtained for other variables (not shown).

5. Interpretation

The basic impacts of stochastic physics on the mean, variance, skewness, and kurtosis of the GFS forecast ensembles can be reproduced in simple one-dimensional systems with additive and multiplicative noise forcing. As discussed in section 3, additive noise generally increases the variance of such systems, but the effects on the mean, skewness, and kurtosis are more subtle. The simplest system in which such effects occur is the linear system B in section 3,

$$dx/dt = -a_0x + (Ex + G)\eta_1 + b\eta_2, \quad (16a)$$

which differs from the even simpler system A only in its inclusion of the extra CAM-noise forcing $(Ex + G)\eta_1$. Here a_0 and b are positive constants, E and G are other constants that may be positive or negative, and η_1 and η_2 are independent Gaussian white noises. The system is probabilistically equivalent to the system considered in (15b),

$$dx/dt = -a_0x + \sqrt{[(Ex + G)^2 + b^2]}\eta, \quad (16b)$$

that is, both are associated with the same Fokker–Planck equation. Figure 3b illustrated how increasing the magnitude of the stochastic forcing in (16b) affects the mean, variance, skewness, and kurtosis of the system.

The particular relevance of system B in interpreting our GFS experiments is that the stochastic perturbations $P(x)r$ in (4) are approximately of the form in (16a). To see this, one may first approximate the anomalous parameterized diabatic tendency as $P' = P - \bar{P} \approx H(x - \bar{x}) + \varepsilon = Hx' + \varepsilon$, where \bar{P} is the time-mean diabatic tendency, H is a regression coefficient, and ε is the regression residual. One may then write

$$\begin{aligned} P(x)r &= (\bar{P} + P')r \approx [Hx + (\bar{P} - H\bar{x})]r + \varepsilon r \\ &\approx (Ex + G)\eta_1 + b\eta_2, \end{aligned} \quad (17)$$

where the red noises r and εr with correlation time scales τ_1 and τ_2 and standard deviations σ_1 and σ_2 are approximated as white noises $r \rightarrow \sigma_1\sqrt{2\tau_1}\eta_1$ and $\varepsilon r \rightarrow \sigma_2\sqrt{2\tau_2}\eta_2$. Defining $E = H\sigma_1\sqrt{2\tau_1}$, $G = (\bar{P} - H\bar{x})\sigma_1\sqrt{2\tau_1}$, and $b = \sigma_2\sqrt{2\tau_2}$ then leads to the stochastic forcing form in (16a).

The differences between the GFS ensemble forecasts with and without stochastic perturbations may thus be crudely interpreted in terms of the differences between systems B and A, both of which are analytically tractable. The change Δx of x in system B over short time intervals Δt may be expressed, using (3), as

$$\begin{aligned} \Delta x &= \left[-\left(a_0 - \frac{1}{2}E^2\right)x + \frac{1}{2}EG \right] \Delta t \\ &\quad + \sqrt{[(Ex + G)^2 + b^2]}\eta\sqrt{\Delta t} \end{aligned} \quad (18)$$

in which the noise-induced drift is seen to reduce the damping rate to $\lambda = a_0 - 0.5E^2$ from $\lambda = a_0$ in system A and also to

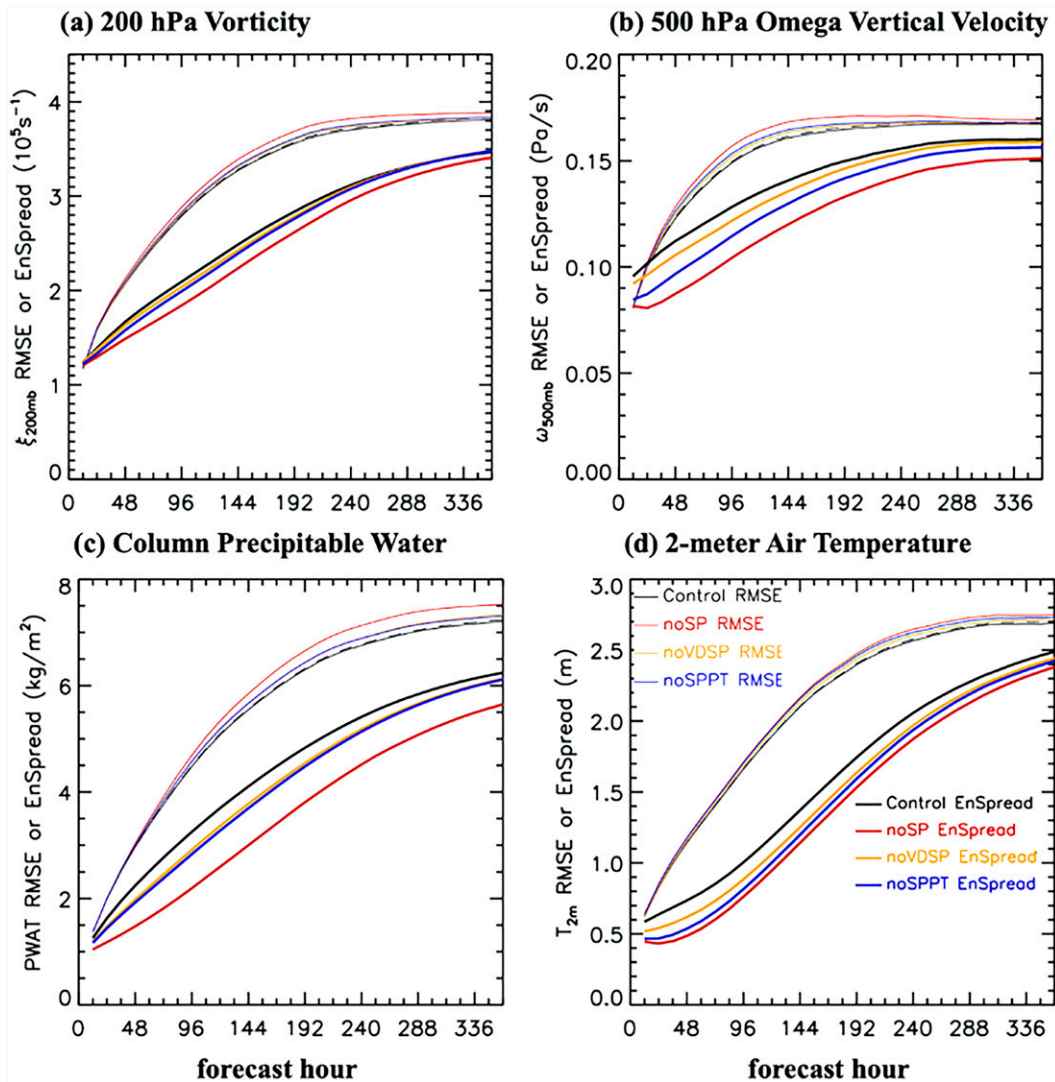


FIG. 12. As in Fig. 1, but showing the impacts on the RMSE and spread from stochastic perturbations to different aspects of the model physics. In each panel, the two groups of four upper and four lower curves respectively show the evolution of RMSE and spread over the 15-day forecast range. The lowest RMSE and highest spread curves, which are identical to the corresponding curves in Fig. 1, show the results for the control forecasts with the full stochastic perturbations. The successively higher RMSE and lower spread curves show the results when the stochastic perturbations to the vertical diffusion tendencies (NO-VDSP), convection and radiation tendencies (NO-SPPT), and the specific humidity (NO-SP) are successively, and cumulatively, switched off. The highest RMSE and lowest spread curves are identical to the corresponding curves with no stochastic perturbations in Fig. 1. An additional dashed curve in all four panels shows the RMSE of the control forecasts if only 40 of the 80 members are used to calculate the ensemble mean.

give rise to an additional constant forcing $0.5Eg$. Both of these effects alter the ensemble-mean forecasts in system B from those in system A as

$$\langle x(t) | x_{in} \rangle = e^{-\lambda t} x_{in} + \frac{EG}{2\lambda} (1 - e^{-\lambda t}). \quad (19)$$

Note that E^2 must be less than $2a_0$ for system B to be stable and yield stationary statistics. Its stationary moments can be derived using the Fokker-Planck equation as

$$\bar{x} = \langle x \rangle = \frac{EG}{2\lambda}, \quad (20a)$$

$$\sigma^2 = \langle (x - \bar{x})^2 \rangle = \frac{(G + E\bar{x})^2 + b^2}{2\lambda - E^2}, \quad (20b)$$

$$S = \frac{\langle (x - \bar{x})^3 \rangle}{\sigma^3} = \frac{2E(G + E\bar{x})}{(\lambda - E^2)\sigma}, \quad \text{and} \quad (20c)$$

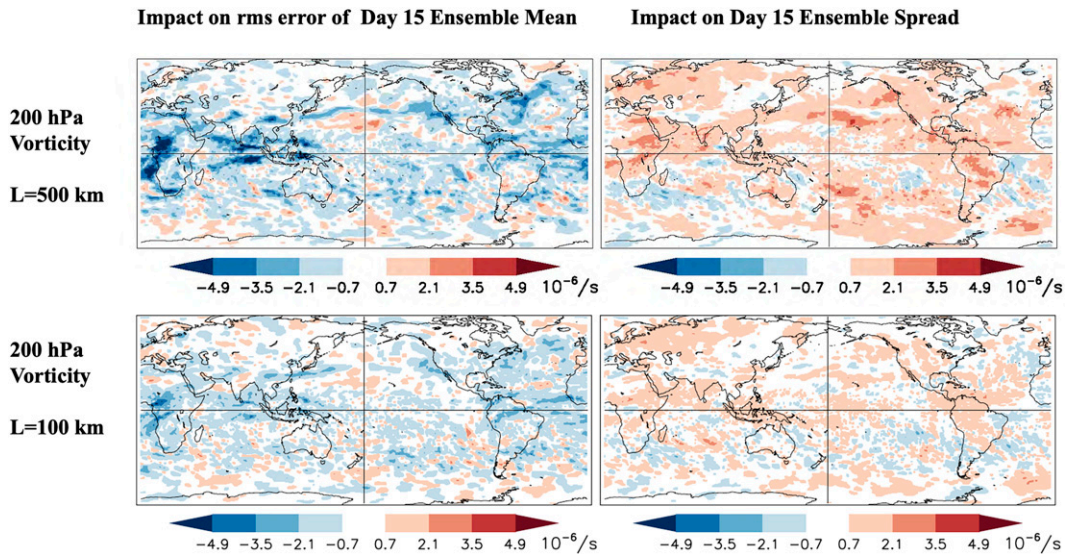


FIG. 13. (top) As in Fig. 6, but for day 15. (bottom) As in the top panels, but for day 15 in otherwise identical forecast ensembles in which the specified correlation length scale of the random perturbations is reduced from 500 to 100 km.

$$K = \frac{\langle (x - \bar{x})^4 \rangle}{\sigma^4} - 3 = \frac{3}{2} \left[\frac{\lambda + E^2}{\lambda - (3/2)E^2} \right] S^2 > \frac{3}{2} S^2. \quad (20d)$$

The stationary moments of system A are obtained by setting E and G equal to 0 in (20). It is then obvious how the additional CAM-noise forcing $(Ex + G)\eta_1$ in system B not only increases the variance but also changes the mean, skewness, and kurtosis. The nondimensional parameter $\tilde{E}^2 = E^2/\lambda$ plays a critical role in determining the magnitude of these changes. Indeed, a necessary condition for the n th statistical moment to exist at all, as evident from the denominators in (20), is $\tilde{E}^2 < 2/(n - 1)$.

Sardeshmukh et al. (2015) have shown that the skewness and excess kurtosis of the probability distributions of large-scale atmospheric anomalies are essentially due to CAM noise forcing. In light of their analysis, the impacts of the CAM-noise like stochastic perturbations in the GFS experiments may be understood even better by considering the impacts of increasing the magnitudes of E , G , and b in system B by a factor $\sqrt{\gamma} > 1$, as done in section 3, than by contrasting system B with system A. The magnitudes of all four moments then increase as illustrated in Fig. 3b, both because of the larger magnitudes of E , G , and b and a further reduction of λ . Additionally, it can be shown using (20) that the change ΔS in S has the same sign as S , and for moderate values of $\sqrt{\gamma} - 1$ is given by $\Delta S = (\sqrt{\gamma} - 1)S$. The change ΔK in K is then given approximately by $\Delta K > 1.5\Delta(S^2) > 1.5(\Delta S)^2$. All of these features are evident in the GFS results for Precip and ω_{500} in Figs. 8 and 9. To a lesser extent, they are also evident in the results for T_{2m} in Fig. 10, although not in the equatorial belt where both S and K are small.

This simple interpretive framework suggests that the multiplicative part of the CAM noise in (17) associated with $E \neq 0$

plays an important role in determining the stochastic impacts on the mean, skewness, and kurtosis of the GFS forecast ensembles. Without E , the forcing in (17) would amount to just increasing the additive noise forcing, a situation considered for the linear system A and nonlinear system C in Fig. 3. However, as already noted in section 3, system A would then predict no change in skewness, and system C would predict decreases in its magnitude, unlike the increases actually obtained in the GFS experiments in Figs. 8–11.

Since $E \neq 0$ in (17) is based on the regression approximation $P' \approx Hx'$, the multiplicative noises in the GFS experiments may be identified with stochasticity in local diabatic feedbacks (positive or negative) on the model variables. Although a detailed discussion is beyond the scope of this paper, the likeliest candidates for such stochastic feedbacks are stochastic mechanical and thermal damping associated with stochasticity in the vertical diffusion of momentum and temperature, and also in cloud-radiative feedbacks, in the model.

To illustrate, for each vertical mode the diffusion amounts to a damping, $P(x) \approx -a_{0v}x$, and its stochasticity can be represented in (16a) with a multiplicative noise parameter $E = -a_{0v}\sigma_1\sqrt{2}\tau_1$ as outlined above. Note that the GFS model also includes an additional horizontal scale-dependent damping as an eighth-order hyperdiffusion that is not stochastically perturbed in our experiments. The total deterministic damping rate for each horizontal wavenumber n may be represented in (16a) as $a_0 = a_{0v} + a_{0h}[n(n + 1)]^4$, where a_{0h} is the hyperdiffusion coefficient divided by the 8th power of Earth's radius. The nondimensional parameter $\tilde{E}^2 = E^2/\lambda$ for each vertical mode and horizontal wavenumber is then

$$\tilde{E}^2 = 2a_{0v}^2\sigma_1^2\tau_1/(a_0 - a_{0v}^2\sigma_1^2\tau_1). \quad (21)$$

This parameter is arguably the single most important determinant of the magnitude of the stochastic impacts on the ensemble-mean, skewness, and kurtosis of the forecast ensembles, not only in our GFS experiments but also generally in other studies that use SPPT types of stochastic perturbations. It explains why the impacts are larger if the amplitude σ_1 and correlation time scale τ_1 of the perturbations r in (4) are increased, and why they are smaller if their spatial correlation scale is decreased, because then the perturbations act preferentially on smaller scales (larger n) for which a_0 in the denominator of (21) is larger. It also explains why specifying large amplitudes and correlation scales of r can cause numerical instabilities in some studies, since this can violate the condition $\tilde{E}^2 < 2$ for the variance to exist [see (20b)].

6. Discussion and concluding remarks

Many previous studies have demonstrated improvement in probabilistic weather forecast skill through increases in forecast ensemble spread generated by randomizing the parameterized diabatic tendencies $P(x)$ as in (4). In this paper we have stressed that such randomizations can also have beneficial impacts on the mean, skewness, and kurtosis of the forecast ensemble distributions throughout the extended 15-day forecast range. We have argued that the impacts on the ensemble-mean forecasts arise, in effect, from noise-induced drifts D in (3) that improve the model by contributing to the deterministic tendency corrections R_d in (1). We have also suggested, through (21) and (18), that to first order the noise-induced drifts effectively reduce the damping of momentum, temperature, and moisture. Such reductions in damping and consequent reduction in system stability also imply increased model sensitivity to external forcing at all time scales.

The idea that randomizing deterministic parameterizations can have beneficial impacts, implying that the details of the parameterizations may not matter beyond some point, can be unpalatable to modelers who often spend years developing them. One reaction has been that although simply “shaking a model” may improve estimates of forecast uncertainty, it may not amount to improving the model itself. We agree, and we have argued that demonstrating reductions in ensemble-mean forecast errors, as done here, is necessary to establish model improvement. Another concern has been that introducing random diabatic perturbations is “unphysical” in that it violates global energy and/or moisture conservation laws. We agree that strict conservation is violated [and possibly also for $P(x)$ itself in many models, although this is rarely discussed]. However, since r is random in both space and time, the global integral of $rP(x)$ is small at individual time steps and even smaller in time averages, so the violation is likely slight. This is evidently the case in our GFS experiments, as illustrated in Fig. 14 for global mean column precipitable water (PWAT). Introducing stochastic perturbations in the model not only does not increase the bias in global mean PWAT but actually decreases it in the first 10 days of the forecasts! In any event, one can argue that violation of strict global conservation constraints is not a fundamental concern, given that strict conservation can always be achieved in principle by expressing the

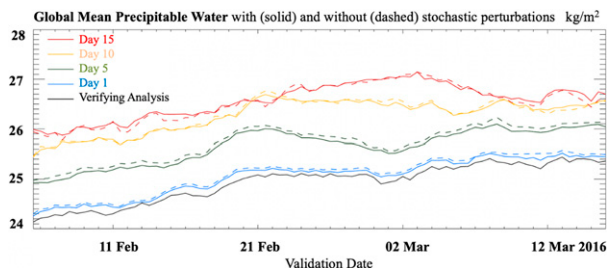


FIG. 14. Ensemble-mean global-mean column precipitable water (PWAT) in the control forecasts with stochastic perturbations (solid curves) and the forecasts without stochastic perturbations (dashed curves). Results are shown for 1-, 5-, 10-, and 15-day forecasts valid at the same calendar dates shown along the abscissa. The lowest curve shows the evolution of PWAT in the verifying analyses over the 6-week period.

diabatic tendencies as convergences of diabatic fluxes and then perturbing the fluxes instead of the tendencies.

We should stress that the beneficial impacts of introducing the $rP(x)$ perturbations are distinct from those obtainable by “improving” $P(x)$ alone. This is because they are associated with two distinct effects of the noise: an increase in forcing (both stochastic and mean, through the noise-induced drift) and a decrease in damping (through the noise-induced drift), as explicitly indicated in (18)–(20) for our simple interpretive model. The two effects cannot be mimicked even in principle by modifying the deterministic parameterizations alone, that is, by changing only the deterministic damping rate a_0 in our interpretive framework.

Stochastic parameterizations are often found to have similar impacts to those obtained by increasing model resolution (e.g., Berner et al. 2012), but are of course much less computationally demanding. While this is good news, it can also be interpreted, wrongly in our view, that stochastic parameterization would become unnecessary at sufficiently high model resolution. Our reason for disagreement is that the specified space and time scales of the random perturbations need to be much larger than the ~ 70 km grid size and ~ 20 min time step of our model to generate substantial beneficial impacts on the ensemble-mean forecasts. A similar remark applies to other stochastic modeling studies as well. Such large-scale perturbations are necessary to account for chaotic physics at even the resolved scales, not just the unresolved scales.

Figure 15 schematically depicts the basic motivation, from our perspective, for introducing fuzziness in model diabatic tendencies. If a model’s parameterized diabatic tendency $P(x, \alpha)$ with parameters α differs from the true diabatic tendency $P_{tr}(x, \alpha_{tr})$, then some correction is obviously necessary. The question is how to proceed if the true tendency is both a rapidly varying function of x and is highly sensitive to the values of α_{tr} . Lacking precise knowledge of either, one possibility is simply to approximate the true tendency as the parameterized model tendency $P(x, \alpha)$ plus random noise with a simple statistical structure, as done in stochastic parameterizations. “Model resolution” does not really enter into this picture. One does need to be mindful if the model physics

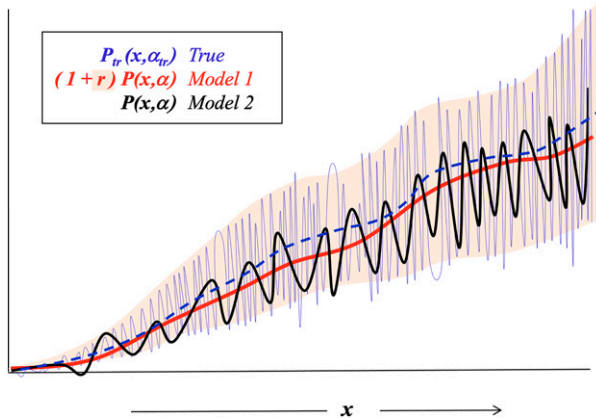


FIG. 15. Basic motivation for stochastic parameterizations of chaotic diabatic tendencies. If the true tendencies P_{tr} (light-blue curve) are both a rapidly varying function of the state variables x and highly sensitive to the precise values of the parameters α_{tr} , they may be approximated in model 1 by augmenting the model's much smoother parameterized tendencies P (red curve) with a stochastic portion $rP(x)$ indicated by the peach swath. If the parameterized tendencies P in another model 2 (black curve) are already somewhat chaotic, they may not need to be augmented by as large of a stochastic swath as in model 1.

are already chaotic to some degree and can statistically represent some portion of the true chaos, in which case one would need to reduce the amplitude of the additional specified stochastic perturbations to avoid the risk of “double counting.” The fact that the spread curve lies below the error curve in almost all forecasting systems is consistent with the idea that the chaotic physics are underrepresented in most models. Additionally, the fact that no stochastic parameterization scheme that we are aware of raises the spread curve above the error curve suggests that the risk of double counting is small.

In the perspective of Fig. 15, the overall uncertainty in $P(x, \alpha)$ is associated with uncertainties in both the functional form of P and the parameters α , that is, with both structural and parametric uncertainties. This overall uncertainty may be crudely accounted for by adding a stochastic portion $rP(x, \alpha)$ to $P(x, \alpha)$ as in the SPPT types of perturbations used in this study. To what extent this can also be accomplished by perturbing only the parameters α as in the alternative stochastic parameter perturbation (SPP) types of perturbation schemes used at some modeling centers (Leutbecher et al. 2017) is still a matter of debate. The two approaches are equivalent for treating the uncertainty in simple linear damping processes, $P(x, \alpha) \approx -\alpha x$, since $rP(x, \alpha) = -r(\alpha x) = -(r\alpha)x$.

A legitimate concern with both the SPPT and SPP types of currently used stochastic perturbation schemes is the ad hoc specification of the amplitudes and space and time scales of the random perturbations r . There is currently no rigorous physical basis for choosing their specific values, and from the perspective of Fig. 15, there will likely never be one, owing to the chaos in the complex interactions among the many different types of diabatic processes in the atmosphere. One rational way forward might be to at least constrain the covariance

structure of the noise empirically, by relating it say to the covariance structure of short-range forecast errors through (1). This is a topic of current research.

We end by stressing that our study in no way diminishes the continuing need to address well-known shortcomings of current deterministic parameterizations of subgrid-scale diabatic processes, as summarized by Stevens and Bony (2013) in their paper “What are climate models missing?”. What it does suggest is that part of what may be “missing” is some accounting of the intractable complexity of rapid diabatic interactions in the atmosphere, that is, the chaotic physics.

Acknowledgments. This work was supported by NOAA/PSL, NOAA/WPO Grant NA19OAR4590158, and the NOAA Cooperative Agreement with CIRES, NA17OAR4320101. Computations were performed on NOAA's High Performance Computing System. The authors gratefully acknowledge PSL's C. McColl for help in managing the large model output and J. S. Whitaker and P. Pegion for help in running the GFS model. Constructive comments by all of the reviewers are also gratefully acknowledged.

Data availability statement. The data used in this research are housed within the NOAA Research and Development High Performance Computing System and are available upon request.

APPENDIX

A Simple Derivation of the Noise-Induced Drift

The noise-induced drift in (3) associated with state-dependent Wiener processes, whose time derivative is state-dependent Gaussian white noise, is a standard result in the theory of stochastic processes (Gardiner 2004; Van Kampen 2007). Here we offer a simple intuitive understanding of its robustness even when the noise is not strictly white. For simplicity we do this in the one-dimensional case. The extension to multiple dimensions is straightforward. Consider the system

$$dx/dt = a(x) + B(x)\eta, \quad (\text{A1})$$

where η is Gaussian red noise $N(0, 1^2)$ with a small correlation time scale τ . The change dx of x over infinitesimal time intervals dt is

$$dx = a(x)dt + B(x)\eta dt. \quad (\text{A2})$$

Now consider the change Δx over finite but still short intervals Δt . Since averages of η over 2τ are nearly independent, one may divide Δt into $k = \Delta t/(2\tau)$ independent noise sub-intervals and treat the sample average $\bar{\eta}$ over Δt as a random variable $N(0, 1^2/k)$. The integral of (A2) from t to $t + \Delta t$ may then be approximated as

$$\Delta x = a(x)\Delta t + \overline{B(x)} \bar{\eta} \Delta t, \quad (\text{A3a})$$

$$= a(x)\Delta t + \overline{B(x)} \phi \sqrt{2\tau} \sqrt{\Delta t}, \quad (\text{A3b})$$

$$= a(x)\Delta t + \overline{\beta(x)} \phi \sqrt{\Delta t}, \quad (\text{A3c})$$

where $\beta(x) = B(x)\sqrt{2\tau}$ and $\phi = N(0, 1^2)$ is a standard Gaussian random variable. Approximating $\beta(x) \approx [\beta(x) + \beta(x + \Delta x)]/2$ (consistent with the “Stratonovich interpretation”) and $\beta(x + \Delta x) \approx \beta(x) + (d\beta/dx)\Delta x$ and rearranging gives

$$\Delta x = \left[a(x)\Delta t + \beta(x)\phi\sqrt{\Delta t} \right] + \frac{1}{2} \frac{d\beta}{dx} \Delta x \phi \sqrt{\Delta t}. \quad (\text{A4})$$

Approximating Δx on the right-hand side as $\Delta x = a(x)\Delta t + \beta(x)\phi\sqrt{\Delta t}$ and rearranging again gives, finally,

$$\Delta x = \left[a(x) + \phi^2 \frac{1}{2} \frac{d\beta(x)}{dx} \beta(x) \right] \Delta t + \beta(x)\phi\sqrt{\Delta t}, \quad (\text{A5})$$

where a third-order term in powers of $\sqrt{\Delta t}$ involving $a(x)(\sqrt{\Delta t})^3$ has been neglected (consistent with “Itô’s lemma”). Equation (A5) is identical to the one-dimensional form of (3), except for the factor ϕ^2 in front of the noise-induced drift inside the square brackets. Although this factor is random, it has a mean of 1 and is always positive, so (A5) always gives the correct sign, and roughly the correct magnitude, of the noise-induced drift in (3).

In a more rigorous derivation, $\overline{B(x)\eta}$ is not approximated as $\overline{B(x)}\overline{\eta}$ in the integral of the noise term in (A3a), and the random factor ϕ^2 in (A5) is replaced by $\overline{\eta^2}$, which is equal to 1 in the limit $k \rightarrow \infty$ of the number k of independent noise subintervals of Δt .

Last, note that since ϕ represents a sample mean over Δt , it can be specified as a different constant number within each interval Δt if the system is marched forward in time, or more smoothly as red noise with a correlation time scale of order Δt . In the atmosphere, one may reasonably expect the validity of (A3a) to extend over intervals Δt comparable to Rossby adjustment time scales (~ 3 h in the extratropics and much longer in the tropics), partly justifying the choice of 6 h for the noise correlation time scale in our GFS experiments.

REFERENCES

- Berner, J., T. Jung, and T. N. Palmer, 2012: Systematic model error: The impact of increased horizontal resolution versus improved stochastic and deterministic parameterizations. *J. Climate*, **25**, 4946–4962, <https://doi.org/10.1175/JCLI-D-11-00297.1>.
- , and Coauthors, 2017: Stochastic parameterization: Toward a new view of weather and climate models. *Bull. Amer. Meteor. Soc.*, **98**, 565–588, <https://doi.org/10.1175/BAMS-D-15-00268.1>.
- Buizza, R., M. Miller, and T. N. Palmer, 1999: Stochastic representation of model uncertainties in the ECMWF ensemble prediction system. *Quart. J. Roy. Meteor. Soc.*, **125**, 2887–2908, <https://doi.org/10.1002/qj.49712556006>.
- Gardiner, C. W., 2004: *Handbook of Stochastic Methods for Physics, Chemistry and the Natural Sciences*. Springer, 415 pp.
- Huffman, D., D. Bolvin, D. Braithwaite, K. Hsu, R. Joyce, and P. Xie, 2014: NASA Global Precipitation Measurement Integrated Multi-satellite Retrievals for GPM (IMERG). Algorithm Theoretical Basis Doc., version 4.4, 26 pp., https://pds.gsfc.nasa.gov/Documents/IMERG_ATBD_V4.pdf.
- Klinker, E., and P. D. Sardeshmukh, 1992: The diagnosis of mechanical dissipation in the atmosphere from large-scale balance requirements. *J. Atmos. Sci.*, **49**, 608–627, [https://doi.org/10.1175/1520-0469\(1992\)049<0608:TDOMDI>2.0.CO;2](https://doi.org/10.1175/1520-0469(1992)049<0608:TDOMDI>2.0.CO;2).
- Leutbecher, M., and Coauthors, 2017: Stochastic representations of model uncertainties at ECMWF: State of the art and future vision. *Quart. J. Roy. Meteor. Soc.*, **143**, 2315–2339, <https://doi.org/10.1002/qj.3094>.
- Palmer, T. N., 2019: Stochastic weather and climate models. *Nat. Rev. Phys.*, **1**, 463–471, <https://doi.org/10.1038/s42254-019-0062-2>.
- , R. Buizza, F. Doblas-Reyes, T. Jung, M. Leutbecher, G. J. Shutts, M. Steinheimer, and A. Weisheimer, 2009: Stochastic parametrization and model uncertainty. ECMWF Tech. Memo. 598, 44 pp., <https://www.ecmwf.int/sites/default/files/elibrary/2009/11577-stochastic-parametrization-and-model-uncertainty.pdf>.
- Penland, C., and P. D. Sardeshmukh, 2022: A mechanism for the skew of ensemble forecasts. *Quart. J. Roy. Meteor. Soc.*, **148**, 1131–1143, <https://doi.org/10.1002/qj.4251>.
- Rodwell, M. J., and T. N. Palmer, 2007: Using numerical weather prediction models to assess climate models. *Quart. J. Roy. Meteor. Soc.*, **133**, 129–146, <https://doi.org/10.1002/qj.23>.
- Sardeshmukh, P. D., 2005: Issues in stochastic parameterization. *ECMWF Workshop on Representation of Sub-Grid Processes Using Stochastic-Dynamic Models*, Reading, United Kingdom, ECMWF, 5–12, <https://www.ecmwf.int/sites/default/files/elibrary/2005/12092-issues-stochastic-parametrization.pdf>.
- , and B. J. Hoskins, 1988: The generation of global rotational flow by steady idealized tropical divergence. *J. Atmos. Sci.*, **45**, 1228–1251, [https://doi.org/10.1175/1520-0469\(1988\)045<1228:TGOGRF>2.0.CO;2](https://doi.org/10.1175/1520-0469(1988)045<1228:TGOGRF>2.0.CO;2).
- , and P. Sura, 2009: Reconciling non-Gaussian climate statistics with linear dynamics. *J. Climate*, **22**, 1193–1207, <https://doi.org/10.1175/2008JCLI2358.1>.
- , and C. Penland, 2015: Understanding the distinctively skewed and heavy tailed character of atmospheric and oceanic probability distributions. *Chaos*, **25**, 036410, <https://doi.org/10.1063/1.4914169>.
- , G. P. Compo, and C. Penland, 2015: Need for caution in interpreting extreme weather statistics. *J. Climate*, **28**, 9166–9187, <https://doi.org/10.1175/JCLI-D-15-0020.1>.
- Stevens, B., and S. Bony, 2013: What are climate models missing? *Science*, **340**, 1053–1054, <https://doi.org/10.1126/science.1237554>.
- Strommen, K., H. M. Christensen, D. MacLeod, S. Juricke, and T. N. Palmer, 2019: Progress towards a probabilistic Earth system model: Examining the impact of stochasticity in the atmosphere and land component of EC-Earth v3.2. *Geosci. Model Dev.*, **12**, 3099–3118, <https://doi.org/10.5194/gmd-12-3099-2019>.
- Van Kampen, N. G., 2007: *Stochastics Processes in Physics and Chemistry*. 3rd ed. Elsevier, 463 pp.
- Wang, J.-W. A., and P. D. Sardeshmukh, 2021: Inconsistent global kinetic energy spectra in reanalyses and models. *J. Atmos. Sci.*, **78**, 2589–2603, <https://doi.org/10.1175/JAS-D-20-0294.1>.
- , —, G. P. Compo, J. S. Whitaker, L. C. Slivinski, C. M. McColl, and P. J. Pegion, 2019: Sensitivities of the NCEP Global Forecast System. *Mon. Wea. Rev.*, **147**, 1237–1256, <https://doi.org/10.1175/MWR-D-18-0239.1>.
- Zhu, Y., and Coauthors, 2018: Toward the improvement of sub-seasonal prediction in the NCEP Global Ensemble Forecast System (GEFS). *J. Geophys. Res. Atmos.*, **123**, 6732–6745, <https://doi.org/10.1029/2018JD028506>.